Measuring Self-Efficacy in Immersive Virtual Learning Environments

Lehikko, Anu

Published in:
Journal of Interactive Learning Research

Published: 15.10.2021

Document Version
Peer reviewed version

Citation for pulished version (APA):
Lehikko, A. (2021). Measuring Self-Efficacy in Immersive Virtual Learning Environments: A Systematic Literature Review. *Journal of Interactive Learning Research*, *32*(2), 125-146.

Journal of Interactive Learning Research (2020) **32**(2), 125-146

Measuring Self-Efficacy in Immersive Virtual Learning Environments: A Systematic Literature Review

ANU LEHIKKO
University of Lapland, Finland
alehikko@ulapland.fi

To further the study of learning in the context of immersive virtual technologies, this systematic literature review addressed the following research questions: 'How is self-efficacy being studied in immersive virtual learning environments?' 'What kinds of self-efficacy findings have been made in immersive virtual learning environments?' The peer-reviewed book chapters and journal articles included in this study were published in English between January 2014 and January 2020.

A narrative synthesis was created for understanding the various aspects of the topic. During the process, the scarcity of experimental research on self-efficacy in immersive virtual learning environments was revealed. The results bring to light the mixed self-efficacy findings in the context of immersive virtual learning and underline the importance of ensuring the quality of the experimental designs in future research. The study of self-efficacy in this context is suggested to benefit from increased attention to the questions of methodology and measurement. To further the goals of immersive virtual learning research, extending the participant base of the experiments by engaging in working life collaboration is strongly recommended.

Keywords: Immersion, virtual learning environments, self-efficacy

Immersive virtual learning environments (VLEs) are considered suitable for the training and development of professional skills in many disciplines, including those in the arts and humanities, health sciences, military and aerospace, and science and technology (Concannon et al., 2019; Pelargos et al., 2017). The technical affordances of immersive virtual reality (VR) have also been found to have potential for teacher training (Billingsley et al., 2019). Advances in computer technology have only recently enabled the development of high-fidelity augmented reality (AR) and VR that can appeal to the masses and capture the interest of commercial enterprises and researchers. During the 2010s, many large corporations released VR and AR systems intended for commercial use, thus making them more easily available for researchers and educators alike (Concannon et al., 2019; Makransky & Lilleholt, 2018; Pelargos et al., 2017).

From education researchers' point of view, immersive technologies may require current cognitive theories of learning to be reassessed from the viewpoint of affective and motivational factors (Makransky, Terkildsen, & Mayer, 2019). Considered a crucial factor associated with educational success in both individual and group contexts, self-efficacy (SE) is a construct that refers to a person's belief in their ability to successfully perform various tasks and to achieve specific goals (Bandura, 1994; Klassen & Usher, 2010). However, previous SE research has been found to suffer from both faulty conceptualisation of the construct and measurement problems. When SE measures are employed in research, their quality should be carefully assessed (Klassen & Usher, 2010).

Through a systematic literature review, the current study aims to contribute to the field of education research by focusing on studies that have utilised SE measures in research on immersive VLEs since the wider availability of the commercial VR technology to the researchers in the mid-2010's (c.f. Lehikko, 2020). The key concepts for this systematic literature review are VLEs, immersion, and SE.

VIRTUAL LEARNING ENVIRONMENTS

Like traditional learning environments, VLEs significantly differ from one another in terms of the types of learning experiences they can offer. A learning management system such as Moodle can be regarded as a type of VLE (Dalgarno & Lee, 2010). When delivered on a digital platform, learning can be either synchronous (instructor-led) or asynchronous (self-paced study). The learning content can be accessed, for example, with a computer or a smartphone (Clark & Mayer, 2016). Three-dimensional (3D) VLEs can be distinguished from other virtual environments by their three-dimensionality, interactivity, and smooth temporal changes (Dalgarno & Lee, 2010).

They often aim to simulate or replicate an environment, sometimes with the aid of VR, AR, or mixed reality (MR) technology (e.g. Makransky & Lilleholt, 2018).

The main difference between VR, AR, and MR lies in the amount of digital content that is mixed with reality (Elbamby et al., 2018). In AR and MR, computer-generated virtual objects are superimposed on the real world and viewed with glasses developed for this specific purpose. VR refers to a completely virtual, simulated experience (Concannon et al., 2019; Elbamby et al., 2018; Pelargos et al., 2017). Representational fidelity, combined with interactional elements, is known to enhance the sense of presence within a virtual environment (Dalgarno & Lee, 2010).

Often, the virtual environment demands the use of an electronic representative called an avatar. Second Life is an example of a widely studied, desktop-based 3D virtual world that belongs to this category (Concannon et al., 2019). VLEs also often incorporate virtual pedagogical agents that are thought to facilitate the learning process. Agents and their design have a significant influence on the learner's SE, which might in turn affect learning outcomes (Schroeder et al., 2018).

IMMERSION

Immersion can be described as 'an objective measure of the extent to which the system presents a vivid virtual environment while shutting out physical reality' (Cummings & Bailenson, 2016, p. 3). The level of immersion separates a desktop virtual environment, which is displayed on a computer screen and is often controlled with a mouse, from an environment that is viewed using a head-mounted display (HMD) or a Cave Automatic Virtual Environment (CAVE), which surrounds the user with large screens and is viewed through a pair of 3D glasses (e.g. Makransky & Lilleholt, 2018).

In some previous studies, the level of immersion has been understood as the user's personal experience of either being convincingly immersed in the virtual world or remaining cognizant of the virtual experience (e.g. Pelargos at al., 2017). The current study reflects the alternative view that a high or low level of immersion in a VLE depends on the technology used to create the experience (Table 1). Desktop VLEs are thus considered non-immersive (Concannon et al., 2019; Makransky & Lilleholt, 2018).

Table 1
Defining Immersive vs. Non-Immersive VR (Concannon et al., 2019, p. 3)

Term	Definition	Examples
Immersive VR	The user is entirely surrounded by the virtual environment, encompassing an optimal field of view (Rebelo et al., 2012).	HMD VR CAVE
Non-immersive VR	The user is not entirely surrounded by the virtual environment; images of the virtual world and the real world can be seen simultaneously (Rebelo et al., 2012).	AR Desktop computer experience

The more immersive the system, the more likely the user will become psychologically engaged in the virtual task at hand (Cummings & Bailenson, 2016). In some users, immersive VR has been found to cause nausea and dizziness, which are similar to the symptoms of motion sickness (e.g. Elbamby et al., 2018). Little empirical research exists on how the level of immersion impacts learners' interest, motivation, self-regulation, and performance in the learning process (e.g. Makransky & Lilleholt, 2018). High fidelity does not necessarily mean a better learning experience; for instance, it may tax the learner's capacity to process sensory data, thus hindering the cognitive processes essential to learning (Makransky, Terkildsen, & Mayer, 2019). From researchers' point of view, high fidelity could be helpful when aiming to measure actual, transferable phenomena instead of phenomena that could be attributed to the environment itself (Billingsley et al., 2019).

SELF-EFFICACY

SE refers to 'people's beliefs about their capabilities to produce designated levels of performance that exercise influence over events that affect their lives' (Bandura, 1994, p. 2). The concept of SE is essential in the social cognitive theory. SE beliefs are especially critical in educational settings and give the learners the impetus to overcome the various challenges often present in situations that demand adaptation to and learning of new skills. The learner's SE should optimally slightly exceed their current ability level for the added motivation to succeed in the task at hand. In general, people prefer activities in which they feel competent over those in which they feel uncertain (Bandura, 2012; Klassen & Usher, 2010). In addition to its direct effect on behaviour, SE influences other determinants such as the level of chosen challenges and goals and the outcome expectations related to these. SE is also reflected in the resilience people demonstrate when working towards these goals (Bandura, 2008; Pajares, 1996).

SE-relevant information is received from four sources identified in research: 1) mastery experiences, 2) vicarious learning, 3) social persuasion, and 4) physiological and affective states. Mastery experiences, such the successful execution of a task, can help learners to build a robust sense of SE. Opportunities to learn from failed efforts are usually required to increase resilience. By the principle of vicarious learning, one's belief in their abilities can be enhanced by observing others succeed in the task. The effect is stronger if learners perceive more similarity between the observed person and themselves (Bandura, 1994, 2008). Note that a person with low SE makes more negative comparisons between the model and themselves, which may cause them to benefit less from the vicarious information (Wilde & Hsu, 2019). Overall, the effect on learners' SE depends on their interpretation of the event. The process through which people select or discard the relevant stimuli is not yet clearly understood (Klassen & Usher, 2010).

Pedagogically, effective SE builders act in a way that enables successful mastery experiences for others. They may act as a model for the desired behaviour, utilise the power of social persuasion by providing positive feedback, and support learners in overcoming negative feelings in the learning process. Encouraging learners to assess their self-improvement instead of comparing themselves to others is also important (Bandura, 2008). In designing SE-enhancing learning situations for immersive VLEs, these actions usually performed by human instructors must be delivered to learners by other means by default. However, owing to the phenomenon's nature, a universal SE solution that uniformly suits all learners is not feasible (Klassen & Usher, 2010).

In their 2010 review of 96 studies on SE, Klassen and Usher noted that 51% of the articles examined used measures that were incongruent with the Bandurian theoretical view on the SE construct. To ensure the theoretical congruence of the measure, SE needs be regarded as a context-specific judgement instead of a trait-like construct: the use of domain-specific SE measurements is recommended (Bandura, 2012; Pajares, 1996). When constructing measures, unipolar SE scales ranging from zero to maximum efficacy should be preferred over bipolar scales (Bandura, 2012). Complex activity domains also demand multidimensional assessments of different types of SE beliefs (Bandura, 2012). Contextual and environmental factors should thus be taken into consideration in SE research.

AIM AND RESEARCH QUESTIONS

The aim of this systematic literature review was to create a narrative synthesis on the research on SE in immersive VLEs (Torgerson et al., 2012). The research questions were 'How is SE being studied in immersive VLEs?' and 'What kinds of SE findings have been made in immersive VLEs?'

METHODS

A systematic literature review can provide an overview of the definitions of a concept and be used to examine how the concept has been utilised in research (e.g. Palsa & Ruokamo, 2015). A systematic review is 'designed to be explicit, transparent and replicable' to overcome the potential problems often associated with the design of traditional literature reviews (Torgerson et al., 2012, p. 217). To identify all relevant research, hand searching of the key journals, citation searching, and checking of the bibliographies of previous studies can be performed as part of the search process (Torgerson et al., 2012).

SELECTION CRITERIA

The publishing range, from January 2014 to January 2020, was chosen owing to the research interest in immersive virtual technologies during this period, which has been growing since the mid-2010s according to Elbamby et al. (2018). Only peer-reviewed journal articles and book chapters published in English were included in the study. Moreover, in accordance with the research questions, only studies in which SE was measured in a learning scenario in an immersive VLE were included. Thus, studies that focused on learning technology not considered as immersive in this study, such as a desktop virtual environment, AR, or MR, were excluded. Uncontrolled experiments were excluded as a quality assurance measure (Sung et al., 2019).

SEARCH PROCEDURE

An initial search was conducted in October 2019 that covered six databases and used the search term 'self-efficacy' or 'self efficacy' combined with one of the following phrases: 'virtual learning', 'learning in virtual environments', 'virtual learning environment', '3D learning environment', '3D learning', or 'immersive learning environment'. The university library informatician was consulted when selecting the suitable databases, search terms, and keywords.

The search covered the following databases: Academic Search Elite, ERIC (ProQuest), PsycINFO, ScienceDirect (Elsevier), Scopus, and Web of Science. These databases were selected for their relevance to educational research. A search was also made in the Finnish databases Melinda and ARTO using the translated search terms and available keywords from the YSO General Finnish Ontology database. The Finnish databases yielded zero results and were excluded from the subsequent search strategy.

The second search was made in December 2019, covering the year range 2014–2019, in the same international databases with the addition of SpringerLink. The search criteria were modified to improve the accuracy of the results. The non-hyphenated search phrase 'self efficacy' was abandoned in favour of the spelling in APA style. The search terms used were 'self-efficacy' in combination with 'immersive', 'immersion', '3D', or 'virtual' as well as the phrase 'learning environment'. The search was updated in January 2020 and July 2020 to obtain results for the full desired year range. The third search was performed in July 2020 with the search terms 'self-efficacy', 'immersive' or 'immersion', and 'virtual reality' or 'VR'. After removing the duplicates, the total number of results retrieved from these searches was 4595.

After a primary body of peer-reviewed literature was identified in each database, the texts were compared against the search terms by reading the titles and abstracts (see Fig. 1). Owing to the multiple interpretations of the relevant glossary present in the field of research, screening according to the level of immersion was not possible at this stage (Concannon et al., 2019). Any publications that did not conform to the inclusion criteria, such as conference papers, were excluded. Only studies that measured general or domain-specific SE were included in the selection (e.g. Shu et al., 2019).

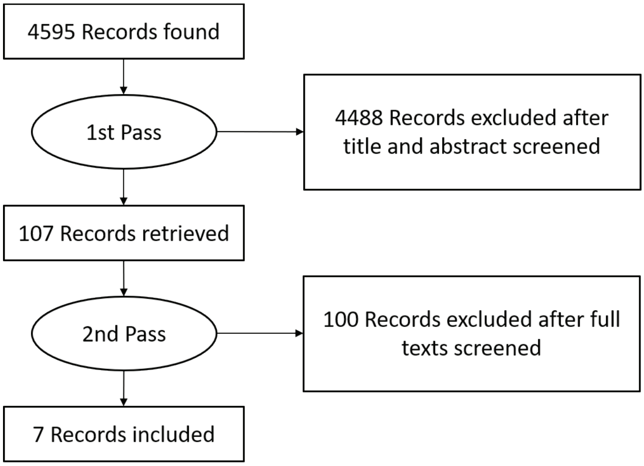

Figure 1. Inclusion and Exclusion Process.

After omitting the duplicates, 107 research articles and book chapters were selected for closer scrutiny. These texts were then read in full. The level of immersion present in the learning environments in each of the studies was assessed based on the glossary proposed by Concannon et al. (2019). Some studies reported the use of immersive VR technology in experiments but did not separately report the results for these technologies (e.g. Farra et al., 2019). Only the studies that demonstrated the use of immersive VLEs in a controlled experiment and reported the results accordingly were selected for the final review; in total, seven studies were included.

In the last search phase, the titles on the lists of references in the selected studies were manually searched for further results. The journals that had published the articles were also searched online using the search terms and criteria described above. No additional literature was found during this phase.

STUDY REVIEW PROCEDURE

First, the selected articles were alphabetically organised in rows by citation on a spreadsheet. The following information was then extracted from each of the articles and organised into columns: the study purpose (e.g. the impact of HMD VR on science SE), the study methodology (e.g. number of comparison groups and treatments and data collection arrangements), the SE measures used (e.g. adapted from the general SE scale or Motivated Strategies for Learning Questionnaire (MSLQ)), the participant features (e.g. age, number, and participant acquisition method), the subject matter and educational context of learning (e.g. aviation safety training), the type of learning media used in the study (e.g. different types of VR displays), and the findings and observations made by the author(s) on learners' SE.

RESULTS

The following variables were summarised from the seven articles that met the pre-determined criteria of measuring SE in an immersive VLE: a) the purpose of the study, b) the learning context, c) the study methodology, d) the sample description, e) the SE measurement, and f) the SE findings (see Table 2). Leder et al. (2019) only measured SE in Study 1, the first of the two experiments described in the paper. Therefore, Leder et al.'s Study 2 was not included in the analysis. Shu et al. (2019) also reported two experiments, which will be referred to here as Shu experiments 'A' and 'B'. The total number of experiments in the summary is eight.

Table 2
Summary of Selected Studies

Study	Study Purpose	Learning Context	Study Methodology	Sample	SE Measurement	SE Findings
Buttussi and Chittaro (2018)	To determine the effects of different types of VR displays on presence and learning in a safety training scenario	Aviation safety training	Controlled experiment, three treatments, pre- and post-tests, retention test, 3 × 3 mixed design ANOVA, Bonferroni post-hoc comparison	Volunteers, n = 96, 32 in each group, aged 18–36 years	Six items on a unipolar seven-point Likert scale, adapted from general SE scale. Measured in pre-test, post-test and retention test	SE significantly increased from the pre-test to the post-test, regardless of the type of display. SE further increased in the retention test.
Huang (2019)	To examine the impact of HDM VR on the science SE of high school students	Science lesson about human cells	Controlled experiment, two treatments, a pre-test on SE for the control group and a post-test on SE for the test group, one-way ANOVA, independent samples t-test	High school 11th grade students, n = 66, 33 in each group, ages not disclosed	Twenty items on a bipolar seven-point Likert scale, science SE, measured either in pre- or post-test	The HMD VR learning activity had no significant effect on the students' science SE.

Table 2 Summary of Selected Studies (**Continued**)

Leder et al. (2019)	To compare learning results, risk perception, and decision-making between CAVE VR and a slideshow presentation	Safety training for the use of pillar drills	Controlled experiment, two treatments, pre- and post-tests, learning outcomes measured and a decision-making task presented in post-test, 2 × 2 mixed design, linear mixed effects analysis	High school students, n = 53, group sizes not disclosed, mean age = 18.42 years (±1.77)	SE measured as a personality measure in pre-test, within the German version of Locus of Control scale	Participants with high SE scores identified more hazards in the VR condition.
Makransky, Borre-Gude, and Mayer (2019)	To analyse the motivational and cognitive benefits of training in immersive VR based on multiple assessments	Laboratory safety training	Controlled experiment, three treatments, pre- and post-test questionnaires, retention test, behavioural transfer tests, one-way between subjects ANCOVA	University students, n = 105, group sizes 35-37-33, majority aged 19–22 years	Seven items on a bipolar five-point Likert scale, adapted from the MSLQ, measured in pre- and post-test	The immersive VR simulation led to significant increases in SE compared with the text booklet condition.
Meyer et al. (2019)	To investigate the effect of pre-training when learning through immersive VR and video	Science lesson about human cells	Controlled experiment, four treatments, pre- and post-test + delayed post-test (1 week), two factorial ANOVA and ANCOVA, independent samples t-test	University students, n = 118, group sizes 31-29-30-28, majority aged 18–25 years, 4.2% over 35 years	Five items on a bipolar five-point Likert scale, adapted from MSLQ, measured in post-test and delayed post-test	Pre-training significantly increased students' SE when learning in immersive VR.

Table 2 Summary of Selected Studies (**Continued**)

Qu et al. (2015)	To examine the effects of social evaluation, vicarious experience, cognitive consistency, and praise on students' beliefs, SE, and anxiety in a VR environment	English class for non-native speakers	Controlled experiment, 2 × 2 within-subjects design with four treatments, questionnaires, physiological measurement data, assessment of speech length, repeated-measures univariate ANOVA, paired-samples t-test	University students, n = 26 exposed to all four conditions, aged 20–30 years	One item on a unipolar 11-point Likert scale measuring SE in belief and experience questionnaire (BEQ), measured in post-tests after each treatment	Once the students witnessed the bystanders' positive attitude to their virtual peers, a positive attitude towards them gave their SE a boost and a negative attitude gave their SE a blow.
Shu et al. (2019), experiments A & B	A comparison of presence and SE between HMDs and desktop computer-facilitated virtual environments	Earthquake education	A: Controlled experiment, single-group repeated-measure design (desktop + HMD VR), pre- and post-tests, questionnaires. B: Controlled, single-group design, pre- and post-tests, only HMD VR. Independent sample t-test, paired sample t-test	University students, n = 96; control, n = 37; group A, n = 39; group B, n = 20; aged 19–33 years	Five items on a bipolar seven-point Likert scale, earthquake preparedness SE, control: measured in pre- and post-test, group A: measured in post-test, group B: measured in pre- and post-test	SE was significantly higher in both the desktop and HMD VR groups compared with the control group. No effect was observed between the treatments in Experiment A.

STUDY PURPOSE

SE was a focal concept in four of the studies (Huang, 2019; Meyer et al., 2019; Qu et al., 2015; Shu et al., 2019). One study aimed to assess a range of motivational and cognitive benefits of training in immersive VR (Makransky, Borre-Gude, & Mayer, 2019). Huang (2019) examined the impact of HDM VR on science SE. Meyer et al. (2019) investigated the effect of pre-training when learning through immersive VR and video. Qu et al. (2015) concentrated on the effects of social evaluation, vicarious experience, cognitive consistency, and praise on students' beliefs, SE, and anxiety.

Presence was the focus of two of the studies (Buttussi & Chittaro, 2018; Shu et al., 2019). Buttussi and Chittaro (2018) aimed to determine the effects of different types of VR displays on presence and learning. Shu et al. (2019) compared presence and SE between an HMD and a desktop VLE. Leder et al. (2019) compared learning results, risk perception, and decision-making between CAVE VR and a slideshow presentation. Five of the studies compared immersive VR with either different types of VR technology or other learning media in their experimental design. The media remained the same between the treatments only in the experiments by Huang (2019) and Qu et al. (2015).

LEARNING CONTEXT

Four of the seven studies used immersive VR as a preparatory vehicle for safety or emergency training. Buttussi and Chittaro's (2018) study utilised a serious game for teaching procedural knowledge about aircraft evacuation. Leder et al. (2019) gave pillar drill safety training to the study participants. Shu et al. (2019) trained their participants in earthquake preparedness, whereas Makransky, Borre-Gude, and Mayer (2019) trained their students on laboratory safety. Qu et al. (2015) utilised VR in an English class for non-native speakers. There were also two science lessons about cells: 'Exploring the cell castle in the human body' (Huang, 2019) and 'The Body VR: The journey inside a cell' (Meyer et al., 2019). The learning objectives in the studies appeared to include both information transmitting and procedural skill building (Clark & Mayer, 2016).

STUDY METHODOLOGY

All the reviewed studies were controlled experiments that each included 2–4 treatments. The use of pre- and post-tests, questionnaires, and other instruments in these experiments was reported. The only post-test decision-making task was arranged by Leder et al. (2019). Qu et al. (2015) reported using physiological measurement data (skin conductance and heart rate) and speech length assessment. Only the experiment by Makransky, Borre-Gude, and Mayer (2019) included a behavioural transfer test; performance differences between the high- and low-level immersion conditions were not identified until this test had been conducted.

The studies displayed a varying degree of rigour in terms of their experimental control. Leder et al. (2019) reported two treatments that only differed with respect to the level of immersion. Makransky, Borre-Gude, and Mayer (2019) trained the study participants with a booklet, non-immersive VR, and immersive VR, always maintaining careful consideration of similar learning content and optimised delivery. In the Shu experiments A and B, the control group received only the questionnaires and no training. Shu experiment A utilised a single-group, repeated-measure design. Under these circumstances, making clear distinctions on which elements contribute to the measured change in learning and SE can be difficult (e.g. Clark & Mayer, 2016).

In Huang's (2019) study, an instrument was designed for measuring science SE. Whereas the control group completed the questionnaire before the learning intervention, the order was reversed for the test group. Note that according to Moriarty (2014), this alone may influence the results of the SE questionnaires. Leder et al. (2019) utilised the SE measure in pre-test only. In the Shu experiment A and the experiments of both Qu et al. and Meyer et al., SE was only measured in post-tests. Qu et al. (2015) reported counterbalancing the order of treatments in their single-group design as a control measure against potential learning, order, or fatigue effects.

SAMPLE DESCRIPTION

Three studies did not disclose the country where the experiments were conducted (Buttussi & Chittaro, 2018; Makransky, Borre-Gude, & Mayer, 2019; Meyer et al., 2019). Of these, both Makransky, Borre-Gude, and Mayer's (2019) and Meyer et al.'s (2019) studies were set in large European universities. The remaining four studies were conducted in China (Huang, 2019), Germany (Leder et al., 2019), the Netherlands (Qu et al., 2015), and Taiwan (Shu et al., 2019).

In four of the seven studies, the participants were university students. The majority were over the age of 18. Buttussi and Chittaro (2018) recruited unpaid volunteers through personal contact. High school students were recruited in two studies: in Leder et al.'s (2019) study, the participants' mean age was 18.42 years, and whereas the age of the students was not disclosed in Huang's (2019) study, the participants were described as 11th graders. The oldest participants were reported in Meyer et al.'s (2019) study, where 'a few' (4.2% of the total number) were older than 35 years, and in the Buttussi and Chittaro's (2018) study, where the age range was 18–36 years. Overall, the research subjects in these experiments were mainly students.

The sample sizes varied between 26 and 118, and the 8 experiments had 560 participants in total. The reported treatment group sizes were between 20 and 37, and the group size data was unavailable for Leder et al.'s (2019) experiment.

SELF-EFFICACY MEASUREMENT

Six of the seven studies measured domain-specific SE (Pajares, 1996). In all six studies, the SE measure used was a 5–11-point Likert scale. The measures varied in terms of the number of items, ranging from 1 to 20. In Qu et al.'s (2015) study, the scale was unipolar with extremes of 'not confident at all' and 'very confident'. Buttussi and Chittaro (2018) utilised a similar scale with 'not at all' and 'very' for the questionnaire items. The scale extremes used by Huang (2019), Makransky, Borre-Gude, and Mayer (2019), and Meyer et al. (2019) were 'strongly disagree' and 'strongly agree'. Shu et al. (2019) similarly used 'totally disagree' and 'totally agree' for the extremes of a bipolar scale.

Four studies reported having adapted their measures from other SE measures such as the MSLQ or the general SE scale (Buttussi & Chittaro, 2018; Leder et al., 2019; Makransky, Borre-Gude, & Mayer, 2019; Meyer et al., 2019). Adopting a different approach to the construct, Leder et al. (2019) included SE in their personality measure set that was given to participants pre-test. For this they utilised the German Locus of Control scale (FKK). Huang's (2019) science SE measure was partially based on several previous science-related SE instruments and partially created for the study in question. In Qu et al.'s (2015) study, the only SE item in the belief and experience questionnaire (BEQ) was based on one that was used in previous research. The Mulilis-Lippa earthquake preparedness scale was used as a reference for the earthquake preparedness SE scale developed for Shu et al.'s (2019) study.

SELF-EFFICACY FINDINGS

In four of the seven studies, both immersive and desktop VR simulations were associated with a significant increase in SE (Buttussi & Chittaro, 2018; Makransky, Borre-Gude, & Mayer, 2019; Meyer et al., 2019; Shu et al., 2019). However, Huang (2019) found no significant science SE effect, and Qu et al. (2015) reported mixed SE results. In Qu et al.'s study, after each participant had observed the virtual bystanders' positive attitude towards their virtual peers in the English class, the positive or negative attitude displayed towards the learner during their turn to speak had a significant impact on the learner's SE. A two-way interaction between self-reported anxiety and SE was also discovered (Qu et al., 2015).

In an aviation safety training scenario, SE significantly increased from the pre-test to the post-test, regardless of the type of display (Buttussi & Chittaro, 2018). Leder et al. (2019) found that the participants with higher SE scores identified more hazards in the pillar drill use safety training, but this was only evident in the CAVE VR condition. In the laboratory safety training, the immersive VR simulation led to significant increases in SE compared with the text booklet condition (Makransky, Borre-Gude, & Mayer, 2019). In the earthquake education, SE was significantly higher in both the desktop and HMD VR groups than in the control group that received no training, and no differences were observed between the non-immersive and immersive treatments (Shu et al., 2019).

When learning biological concepts through immersive VR, pre-training significantly increased students' SE (Meyer et al., 2019). Comparing video and immersive VR, there was evidence of interaction between the learning media and the instructional method, which was also evident in the results regarding SE. The level of immersion on its own did not explain the increase in SE, and the increase could not be attributed to the instructional method of pre-training. The authors concluded that more research is needed on the development of SE when different instructional design methods are combined with different media (Meyer et al., 2019).

DISCUSSION

The aim of this study was to create an overview on the studies measuring SE in immersive VLEs; this was done by producing a narrative synthesis on the results of the systematic literature review. During the search process, the scarcity of research in the area became evident. A key factor in the current study is that SE measures have seldom been used in the context of immersive VLEs so far. Other literature reviews on a specific area of research into immersive VR technologies have made similar observations (e.g. Billingsley et al., 2019).

The diminutive number of studies found in this systematic literature review reflects the results of Concannon et al.'s (2019) literature search, which was performed in 2017 and updated in January 2019. In their review, 13 learning theories were found to be associated with immersive VR in post-secondary education. The social cognitive theory was not among these, and SE was also not mentioned in them (Concannon et al., 2019). Note that several studies were excluded from the current review owing to exclusion criteria related to reporting or experimental control.

The first research question was 'How is SE being studied in immersive VLEs?' The studies included in this review represent a wide range of learning contexts, which echoes the findings of earlier research (Concannon et al., 2019; Pelargos et al., 2017). The majority of the studies included were conducted in a university setting with participants selected among the students. Only one study employed volunteers in its design. Many of the studies involved a comparison between learning media: the immersive VLE remained the same across the treatments in only two studies.

In Leder et al.'s (2019) study, SE was regarded as a personality measure, which is a clear diversion from previous recommendations (Bandura, 2012; Klassen & Usher, 2010; Pajares, 1996). The study utilised the German Locus of Control scale. Most of the selected studies measured domain-specific SE, as advised by Pajares (1996), and their SE measures were either partially or fully based on other previously published instruments. In all these studies, SE was measured in a self-reporting questionnaire using a Likert scale, between 5 and 11 points. A unipolar scale has been previously recommended for SE measures, with some reservations expressed towards the use of a modified Likert scale with a neutral midpoint (Bandura, 2012). Only two of the reviewed studies measured SE with a unipolar scale starting from zero to the maximum level of SE belief. The number of items in the measures greatly varied, falling between 1 and 20. Bandura (2012) advised for measuring the strength of SE across a range of performances within an activity domain instead of that for performance on a specific item.

Billingsley et al. (2019) have suggested that educational research into VR environments could benefit from the development of the methodology and measurements. The results of this study indicate that Billingsley et al.'s recommendation also applies in the study of SE in immersive VLEs. For example, Qu et al. (2015) reported that the attitude of virtual bystanders in a language class scenario had an impact on the participants' SE. In their study, SE was measured with only one questionnaire item; however, the authors also utilised physiological measurement data and a speech length assessment to measure learner anxiety that, according to Bandura (1994, 2012), may influence SE. In the future, exploring the use of physiological measurements to complement self-reported and observational data could perhaps further the study of SE in this context.

The second research question was 'What kinds of SE findings have been made in immersive VLEs?' Four of the seven studies included in the final review reported a significant positive effect on the learners' domain-specific SE. Quality appraisal is an important part of the systematic literature review method: uncontrolled experiments were excluded from the current study as a quality measure (Sung et al., 2019; Torgerson et al., 2012). Keeping in mind the varying level of experimental control in these reviewed studies, the results confirm that immersive VLEs may indeed suit SE-building instructional strategies. Further, carefully designed research is needed to determine which factors influence learner SE and contribute to successful training and development in immersive VLEs.

Whereas media comparisons seem to have captivated the interest of many researchers, other approaches to studying learning in immersive VR should also be encouraged. To create evidence-based, SE-building pedagogy for the training and development of professional skills in immersive VLEs, closely examining SE in the context of professional development is advisable. Cooperation with working life could benefit future research designs, both by allowing access to a more heterogeneous participant base and providing opportunities to study learning transfer between the immersive VR environment and the authentic work conditions. When measuring SE, carefully considering the nature of the construct before choosing and applying the instrument might lead to more consistent findings.

LIMITATIONS OF THIS STUDY

An exhaustive search to cover all the relevant studies is typical for a high-quality systematic review; 'grey' or unpublished literature is included with the intention of avoiding selection bias (Metsämuuronen, 2017; Torgerson et al., 2012). This systematic literature search was limited to peer-reviewed journal articles and book chapters that had been published between January 2014 and January 2020. The databases were chosen for their relevance in the field of education research. Thus, the current study may be subject to some database or publication bias.

The search terms were selected to maximise the initial number of scholarly articles retrieved, under the recognition that a universally accepted glossary for terms related to immersive VLEs does not yet exist. Including studies on simulation may have increased the number of retrieved texts. Furthermore, the analysis of the selected articles and the subsequent narrative synthesis were performed by only one researcher; there was no process in place to verify the coding and categorisation of the items by another expert to ensure the reliability of the analysis, which would undoubtedly benefit a systematic review (Torgerson et al., 2012).

CONCLUSIONS

In light of the results of this systematic literature review, little empirical research exists on SE in immersive VLEs. Immersive technologies appear to suit various learning contexts. They also seem promising from the viewpoint of developing learner SE, but the factors contributing to the SE-building process are not yet clear. Rigorous empirical studies are needed to reveal the most successful instructional strategies and pedagogical models for professional training and development in immersive VLEs: expanding the research scope to working life is recommended. To contribute to the study of motivational and affective domains in the field, it would also be advisable to carefully choose the SE instruments and invest time in developing new research methodology.

References

Bandura, A. (1994). Self-efficacy. In V. S. Ramachaudran (Ed.), *Encyclopedia of human behaviour 4* (pp. 71–81). Academic Press.

Bandura, A. (2008). An agentic perspective on positive psychology. In S. J. Lopez (Ed.), *Positive psychology: Expecting the best in people 3* (pp. 167–196). Praeger.

Bandura, A. (2012). On the functional properties of perceived self-efficacy revisited. *Journal of Management*, *38*(1), 9–44. https://doi.org/10.1177/0149206311410606

Billingsley, G., Smith, S., Smith, S., & Meritt, J. (2019). A systematic literature review of using immersive virtual reality technology in teacher education. *Journal of Interactive Learning Research*, *30*(1), 65–90.

Buttussi, F., & Chittaro, L. (2018). Effects of different types of virtual reality display on presence and learning in a safety training scenario. *IEEE Transactions on Visualization and Computer Graphics*, *24*(2), 1063–1076. https://doi.org/10.1109/TVCG.2017.2653117

Clark, R. C., & Mayer, R. E. (2016). *E-learning and the science of instruction: Proven guidelines for consumers and designers of multimedia learning*. Wiley. https://doi.org/10.1002/9781119239086

Concannon, B. J., Esmail, S., & Roduta Roberts, M. (2019). Head-mounted display virtual reality in post-secondary education and skill training. *Frontiers in Education*, *4*, 1–23. https://doi.org/10.3389/feduc.2019.00080

Cummings, J. J., & Bailenson, J. N. (2016). How immersive is enough? A meta-analysis of the effect of immersive technology on user presence. *Media Psychology*, *19*(2), 272–309.

Dalgarno, B., & Lee, M. J. W. (2010). What are the learning affordances of 3-D virtual environments? *British Journal of Educational Technology*, *41*(1), 10–32. https://doi.org/10.1111/j.1467-8535.2009.01038.x

Elbamby, M. S., Perfecto, C., Bennis, M., & Doppler, K. (2018). Toward low-latency and ultra-reliable virtual reality. *IEEE Network*, *32*(2), 78–84. https://doi.org/10.1109/MNET.2018.1700268

Farra, S., Hodgson, E., Miller, E. T., Timm, N., Brady, W., Gneuhs, M., Ying, J., Hausfeld, J., Cosgrove, E., Simon, A., & Bottomley, M. (2019). Effects of virtual reality simulation on worker emergency evacuation of neonates. *Disaster Medicine and Public Health Preparedness*, *13*(2), 301–308. https://doi.org/10.1017/dmp.2018.58

Huang, W. (2019). Examining the impact of head-mounted display virtual reality on the science self-efficacy of high schoolers. *Interactive Learning Environments.* https://doi.org/10.1080/10494820.2019.1641525

Klassen, R. M., & Usher, E. L. (2010). Self-efficacy in educational settings: Recent research and emerging directions. In S. Karabenick & T. C. Urdan (Eds.), *Decade ahead: Theoretical perspectives on motivation and achievement, Vol. 16 Part A* (pp. 1–34). Emerald Group Publishing.

Leder, J., Horlitz, T., Puschmann, P., Wittstock, V., & Schütz, A. (2019). Comparing immersive virtual reality and powerpoint as methods for delivering safety training: Impacts on risk perception, learning, and decision making. *Safety Science*, *111*, 271–286. https://doi.org/10.1016/j.ssci.2018.07.021

Lehikko, A. (2020). Self-Efficacy in immersive virtual learning environments: A systematic review. In *Proceedings of EdMedia + Innovate Learning* (pp. 904–909). Association for the Advancement of Computing in Education (AACE).

Makransky, G., Borre-Gude, S., & Mayer, R. E. (2019). Motivational and cognitive benefits of training in immersive virtual reality based on multiple assessments. *Journal of Computer Assisted Learning*, *35*, 691–707. https://doi.org/10.1111/jcal.12375

Makransky, G., & Lilleholt, L. (2018). A structural equation modeling investigation of the emotional value of immersive virtual reality in education. *Educational Technology Research and Development*, *66*(5), 1141–1164. https://doi.org/10.1007/s11423-018-9581-2

Makransky, G., Terkildsen, T. S., & Mayer, R. E. (2019). Adding immersive virtual reality to a science lab simulation causes more presence but less learning. *Learning and Instruction*, *60*, 225–236. https://doi.org/10.1016/j.learninstruc.2017.12.007

Metsämuuronen, J. (2017). *Essentials of Research Methods in Human Sciences*. SAGE Publications.

Meyer, O. A., Omdahl, M. K., & Makransky, G. (2019). Investigating the effect of pre-training when learning through immersive virtual reality and video: A media and methods experiment. *Computers & Education*, *140*. https://doi.org/10.1016/j.compedu.2019.103603

Pajares, F. (1996). Self-efficacy beliefs in academic settings. *Review of Educational Research*, *66*(4), 543–578.

Palsa, L., & Ruokamo, H. (2015). Behind the concepts of multiliteracies and media literacy in the renewed Finnish core curriculum: A systematic literature review of peer-reviewed research. *Seminar.net*, *11*(2), 101–119. https://journals.hioa.no/index.php/seminar/article/view/2354

Pelargos, P. E., Nagasawa, D. T., Lagman, C., Tenn, S., Demos, J. V., Lee, S. J., Bui, T. T., Barnette, N. E., Bhatt, N. S., Ung, N., Bari, A., Martin, N. A., & Yang, I. (2017). Utilizing virtual and augmented reality for educational and clinical enhancements in neurosurgery. *Journal of Clinical Neuroscience: Official Journal of the Neurosurgical Society of Australasia*, *35*, 1–4. https://doi.org/10.1016/j.jocn.2016.09.002

Qu, C., Ling, Y., Heynderickx, I., & Brinkman, W. (2015). Virtual bystanders in a language lesson: Examining the effect of social evaluation, vicarious experience, cognitive consistency and praising on students' beliefs, self-efficacy and anxiety in a virtual reality environment. *PLoS ONE*, *10*(4), 1–26. https://doi.org/10.1371/journal.pone.0125279

Rebelo, F., Noriega, P., Duarte, E., & Soares, M. (2012). Using virtual reality to assess user experience. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *54*(6), 964–982. https://doi.org/10.1177/0018720812465006

Schroeder, N. L., Yang, F., Banerjee, T., Romine, W. L., & Craig, S. D. (2018). The influence of learners' perceptions of virtual humans on learning transfer. *Computers & Education*, *126*, 170–182. https://doi.org/10.1016/j.compedu.2018.07.005

Shu, Y., Huang, Y., Chang, S., & Chen, M. (2019). Do virtual reality head-mounted displays make a difference? A comparison of presence and self-efficacy between head-mounted displays and desktop computer-facilitated virtual environments. *Virtual Reality*, *23*(4), 437–446. https://doi.org/10.1007/s10055-018-0376-x

Sung, Y., Lee, H., Yang, J., & Chang, K. (2019). The quality of experimental designs in mobile learning research: A systemic review and self-improvement tool. *Educational Research Review*, *28*. https://doi.org/10.1016/j.edurev.2019.05.001

Torgerson, C., Hall, J., & Light, K. (2012). Systematic reviews. In J. Arthur, M. Waring, R. Coe, & L. V. Hedges (Eds.), *Research methods and methodologies in education* (pp. 217–230). SAGE Publications.

Wilde, N., & Hsu, A. (2019). The influence of general self-efficacy on the interpretation of vicarious experience information within online learning. *International Journal of Educational Technology in Higher Education*, *16*(1), 1–20. https://doi.org/10.1186/s41239-019-0158-x